Exercise in combining distributed data.

Tools:
  - ncview: a visualization tool for netCDF format files
  - mppnccombine: a tool for combining distributed data

Exercise setup:

1) copy and unpack the input data

ssam> cd $CSCRATCH/$USER
ssam> mkdir io
ssam> module use -a /ncrc/home2/fms/local/modulefiles
ssam> module load gcp ncview fre/bronx-1 (if you have not done so already)
ssam> cd io
ssam/io> gcp /lustre/ltfs/scratch/Jeffrey.Durachta/ssam/ssam.io-ex.tgz .
ssam/io> tar xvfz ssam.io-ex.tgz

You should see a set of files 19800101.atmos_month_aer.tileN.nc.00XX where N
runs from 1-6 (indicating the cubed sphere tile face) and XX runs from 00-11
indicating the MPI rank for the face.

Ex1. Try ncview on a distributed data file. For example

ssam/io> ncview 19800101.atmos_month_aer.tile3.nc.0007

Look at various files and fields. The model resolution is "c48" (i.e. 48
points in each direction along the cubed sphere face. One can see that
the parallel decomposition produced a subdomain of (48,4) for each
MPI rank. Hence there were 12 ranks for each face (0-11).

Ex2. Run the combine operation

ssam/io> mppnccombine -M atmos_month_aer.tile3.nc 19800101.atmos_month_aer.tile3.nc.*

Be careful to specify a particular tile number. You cannot combine across tiles.

Note that the tool writes the current memory usage to the screen ending with 50656KB.
The last couple of digits in the memory usage will vary due to the details of how
the data ends up getting mapped to memory pages.

To enhance performance, the combine is capable of working with multiple blocks of data
rather than a single block at a time. This is adjusted by use of the -k option.

The default for k is 1 (i.e. work with 1 block at a time).

Try various blocking factors. The size of this data set limits the usable k value to 12 or less. Note also that the memory usage for this data is approximately of the form

$$Total\_Mem = base + k * 18KB$$

This holds when $k > 2$

Also note that you will have to remove or rename the previous combine output as mppnccombine will not overwrite an existing file (in this case, the .nc file).

Ex3. Try ncview on the combined data. For example.

ssam/io> ncview atmos_month_aer.tile3.nc